Article

CWDAT—An Open-Source Tool for the Visualization and Analysis of Community-Generated Water Quality Data

Annie Gray ^{1,*}, Colin Robertson ² and Rob Feick ³

¹ Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada

² Department of Geography and Environmental Studies, Wilfrid Laurier University, Waterloo, ON N2L 3C5, Canada; crobertson@wlu.ca

³ School of Planning, University of Waterloo, Waterloo, ON N2L 3G1, Canada; rob.feick@uwaterloo.ca

* Correspondence: ma2gray@uwaterloo.ca

Abstract: Citizen science initiatives span a wide range of topics, designs, and research needs. Despite this heterogeneity, there are several common barriers to the uptake and sustainability of citizen science projects and the information they generate. One key barrier often cited in the citizen science literature is data quality. Open-source tools for the analysis, visualization, and reporting of citizen science data hold promise for addressing the challenge of data quality, while providing other benefits such as technical capacity-building, increased user engagement, and reinforcing data sovereignty. We developed an operational citizen science tool called the Community Water Data Analysis Tool (CWDAT)—a R/Shiny-based web application designed for community-based water quality monitoring. Surveys and facilitated user-engagement were conducted among stakeholders during the development of CWDAT. Targeted recruitment was used to gather feedback on the initial CWDAT prototype's interface, features, and potential to support capacity building in the context of community-based water quality monitoring. Fourteen of thirty-two invited individuals (response rate 44%) contributed feedback via a survey or through facilitated interaction with CWDAT, with eight individuals interacting directly with CWDAT. Overall, CWDAT was received favourably. Participants requested updates and modifications such as water quality thresholds and indices that reflected well-known barriers to citizen science initiatives related to data quality assurance and the generation of actionable information. Our findings support calls to engage end-users directly in citizen science tool design and highlight how design can contribute to users' understanding of data quality. Enhanced citizen participation in water resource stewardship facilitated by tools such as CWDAT may provide greater community engagement and acceptance of water resource management and policy-making.

Keywords: citizen science; data quality; web application; water quality; community-based monitoring

Citation: Gray, A.; Robertson, C.; Feick, R. CWDAT—An Open-Source Tool for the Visualization and Analysis of Community-Generated Water Quality Data. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 207. <https://doi.org/10.3390/ijgi10040207>

Academic Editors: Sultan Kocaman and Wolfgang Kainz

Received: 2 February 2021

Accepted: 22 March 2021

Published: 1 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Citizen science (CS) encompasses a wide range of topics and investigations, from ornithology to astronomy to meteorology [1]. Despite this heterogeneity, certain barriers are common to many citizen science initiatives [2]. Differences in training, research priorities/interests, and modes of communicating information that often exist between (and within) CS initiatives and the formal scientific community can limit the degree to which CS initiatives influence decision-making processes [3,4]. Other commonly discussed challenges include volunteer retention [5], the generation of actionable information from raw data [6], data sharing and communication, and overall data quality [7]. Reservations regarding the quality/reliability of citizen-collected and citizen-generated data held by much of the wider scientific community are well documented as a barrier not only to the sustainability of CS initiatives, but to the uptake of citizen-generated data in formal scientific circles [5,7–11]. There is a longstanding discourse in the literature regarding data

quality barriers in citizen science initiatives. Specific data concerns include comparisons of data from different sources [12], differing metadata standards [13], species identifications [14], and factors such as uncertainty, accuracy, bias, and precision [2,15]. Despite these well-known challenges, Fonte et al. (2015) [16] noted an overall dearth of guidance on CS data quality control and quality assurance (QAQC). The ability of citizens and CS initiatives to independently analyze, interpret, and communicate reliable, actionable results from their own high-quality data has been identified as a key challenge for community-based monitoring [17] and the output of reliable, actionable information has been observed as an important driver for citizen science volunteers [5].

One mechanism by which these barriers can be addressed is the development of open-source data analysis and support tools [18]. The development of such statistical and computational resources can support local data management (including quality assurance/quality control—QAQC), reinforce notions of data sovereignty, and promote capacity building in the field of citizen science [14]. Not only do data analysis tools have the potential to address the challenge of data QAQC, but they also offer to other interrelated benefits to CS initiatives and their participants such as: education/knowledge generation, the mobilization of local expertise, capacity building, greater levels of engagement, and increased data sovereignty [19]. Citizen data collectors need easy-to-use tools and interfaces to help them to summarize and visualize their data, assess the quality of their observations, build their understanding of data quality and scope, and to see value in the data they are creating in light of the bigger scientific or regional questions at play. iNaturalist, for example, has a robust user community mechanism where more senior/experienced naturalists correct and verify observations from newer participants, often providing detailed explanation and learning in the process (available at <https://www.inaturalist.org>, accessed on 10 March 2021). Other examples in the field of citizen science include Mackenzie DataStream (available at <https://mackenziedatastream.ca/>, accessed on 10 March 2021); eBird Canada (available at <https://ebird.org/canada/home>, accessed on 10 March 2021); and the CitSci.org website (<https://www.citsci.org/>, accessed on 10 March 2021), which allows citizen science initiatives to register their projects and offers numerous supports such as application programming interfaces. Such tools can catalyze local contextual interpretations which is one of the key benefits of citizen participation and can lead to higher quality information. Many open-source tools have been developed specifically for citizen science purposes, including software architectures, databases, and mobile applications [20].

Open-source data analysis tools and interfaces can provide extra levels of accessibility and transparency by allowing users to view and learn about the operations performed on their data and, when appropriate, to independently modify the software code to fit their needs better [20,21]. This is important for building trust among users who collect data and those rely on them for scientific analysis, enhancing technical capacity within communities [22] and increasing participant engagement [1]. Open-source software is typically available free of charge and restrictive licensing requirements which further enhances tool accessibility and can promote the development of communities of practice around open toolsets and methods [23].

The direct involvement of potential end-users in the design and development processes is critical to the success of any analysis or decision support tool. Ongoing engagement can mitigate issues such as retention and user satisfaction by recognizing the interdependencies between technologies and their intended social contexts [24,25]. Repositories such as Github allow tool developers to make their source code publicly available, potentially leading to code “forking” and increased interoperability [21]. Additionally, the ability to independently visualize, analyze, verify, and communicate citizen science data can support the formation of local policies and solutions, which may then be communicated to the wider scientific community and to government as needed—a benefit of citizen science recognized in Muenich et al. (2016) [26] and Weeser et al. (2018) [27]. The independent development of questions, policies, and answers by citizen scientists can leverage local knowledge and strengthen existing relationships between citizen science

initiatives and scientists by promoting a two-way exchange of information, ideas, and feedback [28]. This is in contrast to the typically unidirectional flow of information from scientists to citizens [29,30], which can accentuate power inequities through a complete reliance on ‘third party’ experts for results.

In this paper, we present a novel R/Shiny based web application, the Community Water Quality Data Analysis Tool (CWDAT), that is designed for citizen science initiatives that focus on community-based water quality monitoring (CBWQM). The remainder of this introduction will present and discuss the research context of community-based water quality monitoring, with a focus on the connections between monitoring challenges and the benefits offered by open-source data analysis tools such as CWDAT.

Conservation and protection of the world’s freshwater resources is a vital global goal enshrined in Sustainable Development Goal 6 on Clean Water and Sanitation which aims to ensure availability and sustainable management of water and sanitation for all. Hydrometric monitoring networks provide vital information on hydrological processes and characteristics such as surface water discharge, watercourse morphology, and water quality [29]. Although regulated national-scale monitoring networks are often the primary sources of hydrometric data, their spatial and temporal coverage can be inadequate when considering local/regional water quality trends and characteristics, due to operational and capacity constraints. Community-based water quality monitoring initiatives can augment regulated monitoring network data by filling the spatial and temporal data gaps and by prioritizing parameters, times, and locations of local interest or concern [31–35].

Many tools have been developed to support the analysis and presentation of data related to water resources, such as AkvaGIS [36,37] and the USEPA’s Water Quality Data Analysis Tool (<https://github.com/USEPA/Water-Quality-Data-Analysis-Tool>, accessed on 10 March 2021). However, most tools are limited in terms of their accessibility (e.g., cost, system requirements, program requirements) and/or are tied to specific protocols for data collection and/or format in terms of the input data for which they are designed. For example, the USEPA Water Quality Data Analysis Tool is designed to work exclusively with the USEPA’s WQP Data Discovery Portal. AkvaGIS, while it accepts data directly from the user, requires field-specific data such as piezometer locations which may not be available/appropriate in the context of community-based water quality monitoring. Overall, tools that aim to monitor, manage, and/or predict natural systems often fail to be adopted and used in the contexts for which they were designed [38,39], a barrier especially relevant to the heterogeneous fields of CS and community-based water quality monitoring. Adapting more general-purpose data-analytic tools for water quality data analysis requires significant technical capacity which may not be possible in many community projects geared toward field data collection. A tool designed through an open-source platform, which can be edited and adapted by end-users and developers alike, thus privileging the respective cultures, contexts, information needs, and preferences of the CBWQM initiatives, holds some promise in addressing such challenges.

To address the needs outlined above, an open-source web application—the Community Water Data Analysis Tool (CWDAT)—was developed as part of a wider project aiming to identify and address barriers to citizen science/CBWQM initiatives and utilization of the data generated by such initiatives (i.e., Global Water Citizenship Project, <http://gwc-gwf.ca>, accessed on 10 March 2021). A prototype version of CWDAT was designed and presented to members of the Canadian community-based water quality monitoring field through a series of surveys and interactive sessions. Based on the feedback received, a second version of CWDAT was developed. The remainder of this paper will elaborate on the prototype’s design, the feedback received from potential end-users, and the consequent developments. Finally, the development of CWDAT is discussed within the overarching context of barriers faced by community-based water quality monitoring initiatives and recommendations for future development and capacity building are provided.

2. Methods

CWDAT is an interactive, open-source web application developed using the R/Shiny framework (R version 4.0.2) [40] and hosted using the open-source version of Shiny Server (<https://rstudio.com/products/shiny/shiny-server/>, accessed on 10 March 2021). As noted in 2017 by Hewitt and Macleod [41], the R/Shiny platform offers such advantages as: low-no cost, suitability on touch devices, ease of development/extension, and potential for scientific innovation, even when compared to other open-source development platforms such as Python and QGIS. The overall goal of CWDAT is to support and enhance CBWQM initiatives by providing a free, user-friendly and customizable tool for independent data validation, visualization, summary, and analysis. CWDAT is neither designed nor intended to replace pre-existing analyses, nor to compete with working relationships already established between citizens and scientists, but rather to complement such connections and to give communities a medium for independent, preliminary interaction with their raw data. An instance of the tool can be accessed through a browser at the following location: <https://spatial.wlu.ca/cwdat/>, accessed on 10 March 2021. The source code and files are freely available through GitHub (<https://github.com/thespatiallabatLaurier/waterquality>, accessed on 10 March 2021). The novelty of CWDAT, relative to other open-source tools in the field of citizen science, lies in its ability to read-in users' data, its standalone nature (no other programs or online portals are required), its ability to statistically compare between data sources, its interactive visualization and reporting capabilities, and its specific focus on the field of community-based water quality monitoring.

The development of CWDAT occurred in three stages (see Figure 1). In the first stage, a prototype version of CWDAT was created based on known barriers to CBWQM in terms of data quality, data visualization, and data communication that were identified from academic and grey literature. Section 2.1 outlines the CWDAT interface, its major features, and the initial design choices. In stage 2, the prototype was presented to members of the CBWQM field through a series of surveys, informal discussions, and interactive tasks. Feedback was solicited on the tool's ease of use, its potential to address barriers faced by CBWQM initiatives, and its potential to generate useful information for CBWQM initiatives (see Section 2.2). Stage 3 centered on incorporating user feedback into a second version of CWDAT (see Sections 3 and 4).

Figure 1. Stages of development for Community Water Data Analysis Tool (CWDAT).

2.1. Prototype Overview and Development

The initial CWDAT prototype included the following sections: Data Upload and Properties; Spatial Visualization; Graphic Visualization; Statistics, and Temporal Coverage Summary. These sections were arranged to facilitate a logical workflow [23] of data visualization and reporting starting with the provision of user-generated water quality data. The Data Upload and Properties component of CWDAT (Figure 2) considered the need for robustness against various naming conventions, dataframe structures, and variables. Additionally, CWDAT is designed to be tolerant of sparse data to recognize that only a subset of variables may be collected at some sites.

CWDAT - The Community Water Data Analysis Tool

About CWDAT Reference Dataset Test Dataset Paired Site Comparisons

Would you like to add your own data or use CWDAT's built-in data?

☐ None

☐ Mine

☒ Built-in

Please select the column containing unique site identifiers

SITE_NO

Please select the column containing date information

DATE_TIME_HEURE

Please select the column containing longitude values

LONGITUDE

Please select the column containing latitude values

LATITUDE

Show 5 entries

Search:

	SITE_NO	DATE_TIME_HEURE	FLAG_MARQUEUR	VALUE_VALEUR	SDL_LDE	MDL_LDM
1	AL07NB0001	1/30/2000 14:15		85.6		
2	AL07NB0001	1/30/2000 14:15	<	0.05		
3	AL07NB0001	1/30/2000 14:15		0.144		
4	AL07NB0001	1/30/2000 14:15		0.0001		
5	AL07NB0001	1/30/2000 14:15		0.046		

Showing 1 to 5 of 173,769 entries

Previous 1 2 3 4 5 ... 34754 Next

Figure 2. Data upload and properties—users may provide their own water quality data or use CWDAT’s built-in data to explore the interface.

CWDAT accepts .csv files or either “long” or “wide” data structures, with the following field requirements: sampling site code, latitude, longitude, and sampling date (additional details and descriptions are provided in Table 1). Data provided in a “long” format require the user to identify which columns contain variable names and variable values. Similarly, data provided in a “wide” format require the user to identify which columns represent specific water quality variables. Once data have been added to CWDAT, the user is asked to identify the requisite columns containing latitude, longitude, date, and site identification code. CWDAT places no limitation on the number of non-requisite columns included in a user-provided .csv file.

The Spatial Visualization page (Figure 3) maps the locations of the monitoring sites. Users may click on a site, select a water quality parameter, and view a scatterplot of the corresponding data. An interactive table with sliders and filters is also provided, for the purpose of outlier identification. Both univariate and bivariate graphing functions are provided by the Graphic Visualization page (Figure 4). Using drop-down menus, users may select the variable(s), months, years, and sites they wish to plot. For reporting purposes, a button is provided for users to download their graphs in .png format. Basic statistical summaries, box plots, and normality testing are offered by the Statistics page (Figure 5) with downloading capabilities. The Temporal Coverage Summary page (Figure 6) was designed to help users identify temporal “gaps” in their data and to report on the volume of data collected. As with previous sections, the user-generated graphs are downloadable.

Table 1. Field requirements for user-provided water quality data.

Required Field	Description	Accepted Data Type(s)
Site identifier	A unique identifier for each discrete location water quality samples were collected or measurements were taken. This can be a name, code, number, or other categorical variable. Multiple observations from a single location should all share the same site identifier.	String, integer, float
Latitude	Latitude coordinate in decimal degrees using the WGS84 system.	Float
Longitude	Longitude coordinate in decimal degrees using the WGS84 system	Float
Date	Date of sample collection. Sample collection time may also be included in this column but is not required.	String, Date, POSIXlt (R)
Indicator/Variable(s)	Water quality indicators (e.g., temperature, pH, etc.). For “long” format data, indicator names will be listed in a single column.

Figure 3. Spatial visualization—users can explore the location, ranges, and values of the data.

Figure 4. Bivariate graphic visualization—users may specify which site(s), variable(s), year(s), and month(s) they wish to visualize.

Figure 5. Statistics—users may generate and download PDF reports and individual graphs.

Figure 6. Temporal coverage summary.

In addition to the visualization a preliminary analysis of a single data set, CWDAT offers users the ability to statistically compare the values of one dataset against another. Conceptually, one dataset would serve as a “reference” (for instance data from a regulated monitoring network) and the second dataset would be community-generated. This capability, offered on the Paired Sites Comparison page, is based on the methodology outlined in Kilgour et al. (2017) [42] and allows users to determine if their community-generated “test” data falls within the normal range of the reference data for corresponding sites. Comparing citizen-generated data to an accepted reference is one way to assess the quality or reliability of CBWQM data. This capability was seen to be important for established community based users to check their data prior to submitting them to a larger project database and also for training purposes where a new participant could compare their observations with historical or regional norms.

2.2. CWDAT Community Feedback

The open-source, dynamic nature of the CWDAT application allows for ongoing development and modifications to support the varied needs and preferences of end-users in the community-based water quality monitoring field. To support such development, thirty-two members of the CBWQM field in Canada were asked to share their insight via an independent survey or via facilitated sessions. The potential participants were contacted due to previous collaboration in one of two ways: (1) previous engagement or association with the Global Water Citizenship research project or (2) participation in a roundtable discussion on community-based monitoring jointly convened by the Gordon Foundation, WWF Canada, and Living Lakes Canada in November 2018.

Two options were offered for participation. Option A entailed independent participation using an online survey. The nine questions of Option A. Option B entailed an interactive, facilitated session using both an online survey and interaction with the tool. Option B questions are provided in Supplementary Document S1. In accordance with the survey questions, Option B participants were provided with a step-by-step instructions document (Supplementary Document S2). An informed consent statement was provided

to participants upon the commencement of either survey, in accordance with the ethics approval granted by the Wilfrid Laurier University Research Ethics Board. Option A was offered in an attempt to maximize the number of participants, by offering an alternative for those who did not have the time or did not wish to participate in the facilitated session. The independent survey (Option A) focused on the roles, motivations, priorities, and barriers experienced by the participants via multiple selection and short and long answer questions. Option B included all survey questions from Option A, in addition to a set of five interactive tasks using CWDAT. As the inclusion of potential end-users through all stages of software development is critical to user retention, user satisfaction, and uptake [24], facilitated sessions with informal discussion encouraged more meaningful reflection and detailed feedback on CWDAT's potential value. The step-by-step instructions and survey questions are provided as Supplementary Data. Table 2 provides a summary of the Option B tasks, the relevant functions of CWDAT, and related discussion topics.

Table 2. Facilitated session tasks and topics.

Task	Purpose	CWDAT Section(s)	Informal Discussion Topics
1	Upload a .csv file of water quality data	Data Upload and Properties	File structures and sizes; metadata; sampling protocols and users' experiences with data handling and storage
2	Identify and explore outlier values	Spatial Visualization and Statistics	Data QAQC; users' methods and needs; outlier detection
3	Visualize the data's temporal scope	Temporal Coverage Summary	Sampling designs; CBWQM initiative organization and resources
4	Graph a subset of data	Graphic Visualization data	Data presentation; viewer and stakeholder preferences and needs data
5	Determine if a subset of test data is within the normal range of a reference baseline	Paired Site Comparisons	Data validation; QAQC; confidence in results; analysis outcomes

Two .csv files containing sample water quality data were provided to participants. The first file was meant to represent data coming from a CBWQM initiative [43], the second to represent data coming from a regulated water quality monitoring network [44]. Upon completion of each task, participants were prompted to reflect via ordinal rankings, multiple selection, and short/long answer questions. Finally, participants were asked to give their general impression of CWDAT and its potential value to the CBWQM field, and to provide commentary and suggestions for improvement based on their interaction with CWDAT.

3. Results

3.1. Response

Cumulatively for surveys A and B, 22 total hits to the survey links were recorded ($n = 22$). Of these, 14 resulted in survey completions. Recalling the initial recruitment of 32 potential participants, approximately 44% of contacted individuals completed a survey. Of the 14 completions, eight participants requested a facilitated session ($n = 8$) and six completed the independent survey ($n = 6$). Participants' self-declared roles and motivations (multiple select) were primarily scientific research, environmental awareness, and policy and decision-making (Table 3).

Table 3. Participant roles and motivations.

Role	Count	%
Scientist or researcher	9	64
NGO/Not-for-profit	3	21
Outreach	3	21
Data analyst	2	14
Volunteer	2	14
Government or leadership	1	7
Environmental consulting	1	7
Community member	1	7
Motivation	Count	%
Environmental awareness	7	50
Scientific research	6	43
Planning and decision-making	4	29

3.2. CWDAT Reception

At the end of Option B, users were asked to rank their overall impression of CWDAT based on three criteria: intuitiveness of the interface; relevance to the users' CBWQM data questions; and generation of actionable information, on a scale from 1 (worst)–5 (best). The respective modes were 4, 5, and 5 ($n = 8$). Most participants emphasized the need for tools such as CWDAT, and many expressed an interest in following the tool's development.

Through informal discussion and interaction with CWDAT, the Option B participants of this study outlined and expanded on numerous barriers they, and their respective CBWQM/citizen science initiatives, have faced. Some information was solicited in response to participant commentary on the tool and its features. Other information was volunteered by the participants when describing their experience, future hopes of the field, and procedures in their respective community/organization. Highlighted barriers ranged from initiative-specific challenges to perceived and actual characteristics of CBWQM and citizen science fields. Three general categories of barriers were observed from the transcribed feedback as shown in Table 4: metadata standards, data interpretation, and communication/information sharing. Multiple participants affirmed that the CWDAT prototype could be beneficial to the CBWQM field, while stressing the need for ongoing engagement and development.

Table 4. Summary of recurring identified and discussed barriers.

Metadata Standards	Data Interpretation	Communication/Sharing
Controlling for units	No consistent idea of how to use data	Privacy concerns
Inconsistent data labelling variations in instrumentation and laboratory procedures Variations in naming conventions Variations in file format and data structures	Establishing trends and triggers	Internet capacity
	Long-term analysis capacity	Communication media
		Perceived lack of quality
	Lack of meaningful interpretation, coordination, and common reporting within and between community-based water quality monitoring/citizen scientist initiatives	File sizes

Participant responses to the call for suggestions/next steps for CWDAT included better supplementary information (i.e., explanatory text regarding water quality parameters and plain-language descriptions of the analysis done on the data), enhancement of raw

data sharing capabilities, and future engagement with developing initiatives prior to the completion of a publicly available model. Participants' preferred output media included plain-text summaries, graphs, reports, and maps using colour to spatially display water quality parameters, their values, and associated criteria. One participant connected the need for informal, explanatory text to differences between grassroots community members and the wider scientific community, highlighting that the interface must not be too "intimidating". Discussions with other participants placed the same concern in terms of default templates and settings—advanced users may find default settings restrictive, but too many options and settings could overwhelm and deter users less comfortable with technology [24].

3.3. CWDAT Development

In response to the feedback of participants, particularly those who selected Option B and engaged directly with the CWDAT prototype, several changes were made to the CWDAT interface and features. Major additions included a visual theme for the user interface; built-in sample data; the generation of downloadable, editable PDF reports; and plain language descriptions and explanations. Figure 7 shows CWDAT's initial Data Upload and Properties page. Figure 8 shows the same page following participant feedback.

Water Quality Data Analysis Tool - PRE-ALPHA
Reference Dataset ▼
Test Dataset ▼
Paired Site Comparisons ▼

Please Upload Reference Data CSV File

Browse...
cbmgrab.csv

Upload complete

Please select the column containing unique site identifiers

Site_Code ▼

Please select the column containing date information

Date ▼

Please select the column containing longitude values

Longitude ▼

Please select the column containing latitude values

Latitude ▼

Please select the format of your data

☒ Wide
☐ Long

Show 5 entries
Search:

	Site_Code	Site_Name	Region	Latitude	Longitude	Date
1	STH-SR	Fort Smith at Slave River / Rapids	S. Slave	60.0207	-111.885	6/13/2012
2	RES-SR	Fort Resolution at Slave River / Big Eddy	S. Slave	61.2674	-113.4161	6/13/2012
3	FGH-RMPT	Fort Good Hope at Mackenzie River / Ramparts	Sahtu	66.1601	-128.9561	7/17/2012
4	NW-MR-U/S	Norman Wells at Mackenzie River / Upstream of Imperial Island 1	Sahtu	65.2755	-126.8666	7/17/2012
5	NW-D/S	Norman Wells at Mackenzie River / Downstream at Radar	Sahtu	65.3033	-127.0584	7/17/2012

Figure 7. Data upload and properties (initial interface).

Figure 8. Data upload and properties (following participant feedback).

4. Discussion

4.1. Response and Reception

Although the response rate of 44% was higher than expected, the low overall number of participants, particularly those who interacted with CWDAT via a facilitated session ($n = 8$) substantially limits any claim of CWDAT's value to the community-based water quality monitoring field. Additionally, a better representation of community members and other grassroots stakeholders would enhance the results and give a truer accounting of CWDAT's potential use within the wider CBWQM field. However, the feedback and discussions described below did provide some insight into the barriers faced by CBWQM initiatives and the ability of CWDAT to address certain needs in the field.

4.2. Prototype Modifications and Implications

The facilitated sessions, while limited to a low number of participants, allowed for in-depth discussions and maximized the insight offered by each participant. Moreover, the provision of a working prototype served as a catalyst for more detailed discussions—both in terms of CWDAT's individual development and its value within the broader context of CBWQM and CS. This was critical as it expanded discussion from more general and abstract concerns and interests focused largely on questions of “what” to address both the “what” and the “how” (interface) [45].

As expected, the workflow from raw data to actionable information, and data quality concerns, are two substantial barriers to sustainable community-based water quality monitoring. The heterogeneous nature of the field, as represented by participants in terms of a dearth of consistency in protocols, reporting, and workflows, is another challenge, a

finding consistent with Jollymore et al. (2017) [30] regarding citizen participation in the hydrological sciences.

Field-specific barriers such as water sample data quality must be viewed in the context of initiative-specific barriers and restrictions. For example, the use of laboratory testing, while it can increase the perceived reliability of the data, can create another barrier if consistent laboratory protocols are not used within/across initiatives. If a set of laboratory protocols are established for water quality sample analysis across the field of CBM, it must be considered if all initiatives have the capacity, financial or otherwise, to adhere to such protocols. Participants indicated that the proposed method of statistical paired site comparison is a promising technique which could help to address the discussed barriers. Specifically, the reliance on publicly available datasets can leverage spatial open government data to the benefit of the CBM field, especially as this resource is typically underused outside of the scope of “expert” research projects [46] while remaining accessible.

The provision of the tool’s source code, the literature source for the statistics [42], and requested plain language explanatory text within the tool’s interface speak to transparency and, where desired, community data sovereignty. Transparency is a guiding pillar of web tool development within the CBM field for enhancing watershed management and planning [47]. Data sovereignty recognizes that some communities (e.g., Citizen groups, Indigenous communities) may want to explore and validate their own monitoring data, yet not share their data with an external citizen science project, government, or industry [48,49]. Further development of data QA/QC functions for the tool, as requested by Survey B participants, included the use of colours to flag extraneous values, reflecting a documented characteristic of Decision Support Systems—the identification of conflicting data [50], which participants connected to the challenge of establishing norms and trends across and within monitoring jurisdictions. The potential for such information (via CWDAT) to help improve the consistency of CBM practices was discussed by some participants in terms of temporal and spatial biases in the data—a line of inquiry consistent with Geldmann et al. (2016) [51], which indicated that modelling the intensity (interpreted in this context as “number”) of observations can help to understand spatial and temporal biases at/between monitoring stations.

As discussed at length by one Survey B respondent, while many CBM initiatives have established effective and beneficial working relationships with scientists and formal institutions, the proposed tool has the potential to fill the niche between the grassroots and the highly scientific and technical. By not only allowing users to ask questions of their data but also introducing users to potential questions they had not considered, by virtue of an open-ended design, members of the CBWQM field can be in a better position to understand and leverage their own data (starting small) either independently or in preparation for collaboration. Such discussions aligned well with previously established barriers and best practices in the literature. The use of the open-source R/Shiny framework supports a versatile, open-ended design affirmed by the Option B participants, as opposed to more traditional, “top-down” tool designs. This progression is consistent with findings in Castillo et al. (2016) [52], which suggested future work on Environmental Decision Support Systems will focus on broadening the range of EDSS capabilities and applicability. Although CWDAT should not be considered a full EDSS, the drive toward better features and wider relevance is shared.

The study revealed how obtaining relevant feedback on new software tools in a citizen science context is necessarily time-consuming and application-specific. Thus, identifying generic principles of geospatial capacity building in citizen science initiatives is challenging. However, we found that much of the rich feedback from Survey B respondents, facilitated stronger relationships with the project which are important cornerstones of project sustainability. The technical dimensions of interface design—while important—may be of less overall long-term value than the social dimensions conveyed through the use of the technology as a boundary object between citizen and scientist and/or technologist [53].

5. Conclusions

This paper presented the Community Water Data Analysis Tool, an open-source web application using the R/Shiny platform. CWDAT is intended to support citizen science initiatives in the field of water quality monitoring, especially community-based initiatives. CWDAT's interface allows a user to provide their own water quality data in .csv format and is robust against varying data structures (i.e., long vs. wide), date and time formats, and naming conventions.

A series of facilitated sessions with members of the community-based water quality monitoring field yielded positive feedback for CWDAT, insight into the challenges faced by CBWQM initiatives, and suggestions for future iterations of CWDAT. Feedback on CWDAT was positive and addressed a gap between citizen scientists and the wider scientific community by providing an accessible tool for independent visualization, analysis, and reporting of community-generated water quality data. CWDAT's use of an open-source language (R) with a robust online support community, combined with the provision of CWDAT's source code through Github, allows CBWQM and CS initiatives to modify CWDAT as they see fit. Future iterations of CWDAT will incorporate water quality thresholds and guidelines, the calculation of the Canadian Council of Ministers of the Environment water quality index, and other methods of data presentation and analysis. Overall, feedback from the study participants identified barriers to citizen science initiatives such as data quality and contextual divides between citizens and scientists.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/ijgi10040207/s1>, Document S1: Survey Questions and Document S2: Instructions.

Author Contributions: Conceptualization, Annie Gray and Colin Robertson; methodology, Annie Gray and Colin Robertson; software, Annie Gray; validation, Annie Gray and Colin Robertson; writing—original draft preparation, Annie Gray, Colin Robertson, and Rob Feick; writing—review and editing, Annie Gray, Colin Robertson, and Rob Feick; visualization, Annie Gray; supervision, Colin Robertson; project administration, Colin Robertson; funding acquisition, Colin Robertson. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Canada First Research Excellent Fund [Global Water Futures Project].

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Research Ethics Board of Wilfrid Laurier University (protocol code 5987, approved 14 February 2019).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the participants to publish this paper.

Data Availability Statement: The participant survey data presented in this study are not available for release due to ethical and privacy considerations governed by our research ethics review. Water quality data presented in this study are publicly accessible via the data portal Mackenzie Data Stream located at <https://mackenziedatastream.ca/> and at open.canada.ca.

Acknowledgments: The authors gratefully acknowledge the support of the Gordon Foundation and of members of the community-based water quality monitoring field who took the time to share their insight.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Haklay, M. Citizen Science and Volunteered Geographic Information: Overview and Typology of Participation. In *Crowdsourcing Geographic Knowledge*; Sui, D., Elwood, S., Goodchild, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 105–122.
2. Kosmala, M.; Wiggins, A.; Swanson, A.; Simmons, B. Assessing data quality in citizen science. *Front. Ecol. Environ.* **2016**, *14*, 551–560. [[CrossRef](#)]
3. Jordan, R.C.; Gray, S.A.; Howe, D.V.; Brooks, W.R.; Ehrenfeld, J.G. Knowledge Gain and Behavioral Change in Citizen-Science Programs. *Conserv. Biol.* **2011**, *25*, 1148–1154. [[CrossRef](#)]

4. Roy, H.E.; Pocock, M.J.O.; Preston, C.D.; Roy, D.B.; Savage, J.; Tweddle, J.C.; Robinson, L.D. *Understanding Citizen Science & Environmental Monitoring*; Final Report on behalf of UK-EOF; CEH: Lancaster, UK, 2012.
5. Alender, B. Understanding volunteer motivations to participate in citizen science projects: A deeper look at water quality monitoring. *J. Sci. Commun.* **2016**, *15*, A04. [\[CrossRef\]](#)
6. Carlson, T.; Cohen, A. Linking community-based monitoring to water policy: Perceptions of citizen scientists. *J. Environ. Manag.* **2018**, *219*, 168–177. [\[CrossRef\]](#)
7. Bird, T.J.; Bates, A.E.; Lefcheck, J.S.; Hill, N.A.; Thomson, R.J.; Edgar, G.J.; Stuart-Smith, R.D.; Wotherspoon, S.; Krkosek, M.; Stuart-Smith, J.F.; et al. Statistical solutions for error and bias in global citizen science datasets. *Biol. Conserv.* **2014**, *173*, 144–154. [\[CrossRef\]](#)
8. Bonter, D.N.; Cooper, C.B. Data validation in citizen science: A case study from Project FeederWatch. *Front. Ecol. Environ.* **2012**, *10*, 305–307. [\[CrossRef\]](#)
9. Foody, G.M.; See, L.; Fritz, S.; Van Der Velde, M.; Perger, C.; Schill, C.; Boyd, D.S. Assessing the Accuracy of Volunteered Geographic Information arising from Multiple Contributors to an Internet Based Collaborative Project. *Trans. GIS* **2013**, *17*, 847–860. [\[CrossRef\]](#)
10. Goodchild, M.F.; Li, L. Assuring the quality of volunteered geographic information. *Spat. Stat.* **2012**, *1*, 110–120. [\[CrossRef\]](#)
11. Hunter, J.; Alabri, A.; Van Ingen, C. Assessing the quality and trustworthiness of citizen science data. *Concurr. Comput. Pract. Exp.* **2013**, *25*, 454–466. [\[CrossRef\]](#)
12. Huang, G.H.; Xia, J. Barriers to sustainable water-quality management. *J. Environ. Manag.* **2001**, *61*, 1–23. [\[CrossRef\]](#)
13. Connors, J.P.; Lei, S.; Kelly, M. Citizen Science in the Age of Neogeography: Utilizing Volunteered Geographic Information for Environmental Monitoring. *Ann. Assoc. Am. Geogr.* **2012**, *102*, 1267–1289. [\[CrossRef\]](#)
14. MacPhail, V.J.; Colla, S.R. Power of the people: A review of citizen science programs for conservation. *Biol. Conserv.* **2020**, *249*, 108739. [\[CrossRef\]](#)
15. Senaratne, H.; Mobasher, A.; Ali, A.L.; Capineri, C.; Haklay, M. A review of volunteered geographic information quality assessment methods. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 139–167. [\[CrossRef\]](#)
16. Fonte, C.C.; Bastin, L.; Foody, G.; Kellenberger, T.; Kerle, N.; Mooney, P.; Olteanu-Raimond, A.-M.; See, L. VGI Quality Control. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *2*, 317–324. [\[CrossRef\]](#)
17. Conrad, C.C.; Hilchey, K.G. A review of citizen science and community-based environmental monitoring: Issues and opportunities. *Environ. Monit. Assess.* **2011**, *176*, 273–291. [\[CrossRef\]](#)
18. Bonney, R.; Shirk, J.L.; Phillips, T.B.; Wiggins, A.; Ballard, H.L.; Miller-Rushing, A.J.; Parrish, J.K. Next steps for citizen science. *Science* **2014**, *343*, 1436–1437. [\[CrossRef\]](#)
19. Yadav, P.; Darlington, J. Design Guidelines for the User-Centred Collaborative Citizen Science Platforms. *Hum. Comput.* **2016**, *3*. [\[CrossRef\]](#)
20. de Reyna, M.A.; Simoes, J. Empowering citizen science through free and open source GIS. *Open Geospat. Data Softw. Stand.* **2016**, *1*, 1. [\[CrossRef\]](#)
21. Luna, S.; Gold, M.; Albert, A.; Ceccaroni, L.; Claramunt, B.; Danylo, O.; Haklay, M.; Kottmann, R.; Kyba, C.; Piera, J.; et al. Developing Mobile Applications for Environmental and Biodiversity Citizen Science: Considerations and Recommendations. In *Multimedia Tools and Applications for Environmental & Biodiversity Informatics*; Springer International Publishing: Cham, Switzerland, 2018; pp. 9–30.
22. Fernandez-Gimenez, M.E.; Ballard, H.L.; Sturtevant, V.E. Adaptive Management and Social Learning in Collaborative and Community-Based Monitoring: A Study of Five Community-Based Forestry Organizations in the western USA. *Ecol. Soc.* **2008**, *13*, 4. Available online: <http://www.ecologyandsociety.org/vol13/iss2/art4/> (accessed on 10 March 2021). [\[CrossRef\]](#)
23. Brenton, P.; von Gavel, S.; Vogel, E.; Lecoq, M.E. Technology Infrastructure for Citizen Science. In *Citizen Science: Innovation in OpenScience, Society and Policy*, 1st ed.; Hecker, S., Haklay, M., Bowser, A., Makuch, Z., Vogel, J., Bonn, A., Eds.; UCL Press: London, UK, 2018; pp. 63–80.
24. Skarlatidou, A.; Hamilton, A.; Vitos, M.; Haklay, M. What do volunteers want from citizen science technologies? A systematic literature review and best practice guidelines. *J. Sci. Commun.* **2019**, *18*, A02. [\[CrossRef\]](#)
25. Klein, L. What do we actually mean by ‘sociotechnical’? On values, boundaries and the problems of language. *Appl. Ergon.* **2014**, *45*, 137–142. [\[CrossRef\]](#)
26. Muenich, R.; Peel, S.; Bowling, L.; Haas, M.; Turco, R.; Frankenberger, J.; Chaubey, I. The Wabash Sampling Blitz: A Study on the Effectiveness of Citizen Science. *Citiz. Sci. Theory Pract.* **2016**, *1*, pe0188507. [\[CrossRef\]](#)
27. Weeser, B.; Kroese, J.S.; Jacobs, S.R.; Njue, N.; Kemboi, Z.; Ran, A.; Breuer, L. Citizen science pioneers in Kenya—A crowdsourced approach for hydrological monitoring. *Sci. Total Environ.* **2018**, *631–632*, 1590–1599. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Capdevila, A.S.L.; Kokimova, A.; Ray, S.S.; Avellán, T.; Kim, J.; Kirschke, S. Success factors for citizen science projects in water quality monitoring. *Sci. Total Environ.* **2020**, *728*. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Hall, G.B.; Chipeniuk, R.; Feick, R.D.; Leahy, M.G.; Deparday, V. Community-based production of geographic information using open source software and Web 2.0. *Int. J. Geogr. Inf. Sci.* **2010**, *24*, 761–781. [\[CrossRef\]](#)
30. Jollymore, A.; Haines, M.J.; Satterfield, T.; Johnson, M.S. Citizen science for water quality monitoring: Data implications of citizen perspectives. *J. Environ. Manag.* **2017**, *200*, 456–467. [\[CrossRef\]](#)
31. Keum, J.; Kaluarachchi, J.J. Development of a decision-making methodology to design a water quality monitoring network. *Environ. Monit. Assess.* **2015**, *187*, 1–14. [\[CrossRef\]](#) [\[PubMed\]](#)

32. Hadj-Hammou, J.; Loisel, S.; Ophof, D.; Thornhill, I. Getting the full picture: Assessing the complementarity of citizen science and agency monitoring data. *PLoS ONE* **2017**, *12*, e0188507. [CrossRef]
33. Penny, D.; Williams, G.; Gillespie, J.; Khem, R. 'Here be dragons': Integrating scientific data and place-based observation for environmental management. *Appl. Geogr.* **2016**, *73*, 38–46. [CrossRef]
34. Walker, D.; Forsythe, N.; Parkin, G.; Gowing, J. Filling the observational void: Scientific value and quantitative validation of hydrometeorological data from a community-based monitoring programme. *J. Hydrol.* **2016**, *538*, 713–725. [CrossRef]
35. Werts, J.D.; Mikhailova, E.A.; Post, C.J.; Sharp, J.L. An Integrated WebGIS Framework for Volunteered Geographic Information and Social Media in Soil and Water Conservation. *Environ. Manag.* **2012**, *49*, 816–832. [CrossRef]
36. Criollo, R.; Velasco, V.; Nardi, A.; de Vries, L.M.; Riera, C.; Scheiber, L.; Jurado, A.; Brouyère, S.; Pujades, E.; Rossetto, R.; et al. AkvaGIS: An open source tool for water quantity and quality management. *Comput. Geosci.* **2019**, *127*, 123–132. [CrossRef]
37. Perdiki, M.; Manjarrez, R.C.; Pouliaris, C.; Rossetto, R.; Kallioras, A. Free and open-source GIS-integrated hydrogeological analysis tool: An application for coastal aquifer systems. *Environ. Earth Sci.* **2020**, *79*, 1–16. [CrossRef]
38. Matthies, M.; Giupponi, C.; Ostendorf, B. Environmental decision support systems: Current issues, methods and tools. *Environ. Model. Softw.* **2007**, *22*, 123–127. [CrossRef]
39. Rodela, R.; Pérez-Soba, M.; Bregt, A.; Verweij, P. Spatial decision support systems: Exploring differences in pilot-testing with students vs. professionals. *Comput. Environ. Urban Syst.* **2018**, *72*, 204–211. [CrossRef]
40. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020. Available online: <https://www.R-project.org/> (accessed on 10 March 2021).
41. Hewitt, R.; Macleod, C. What Do Users Really Need? Participatory Development of Decision Support Tools for Environmental Management Based on Outcomes. *Environments* **2017**, *4*, 88. [CrossRef]
42. Kilgour, B.W.; Somers, K.M.; Barrett, T.J.; Munkittrick, K.R.; Francis, A.P. Testing Against Normal with Environmental Data. *Integr. Environ. Assess. Manag.* **2017**, *13*, 188–197. [CrossRef] [PubMed]
43. Government of Northwest Territories. *NWT-Wide Community-Based Water Quality Monitoring, Environment and Natural Resources*; Government of the Northwest Territories: Yellowknife, NT, Canada, 2019. [CrossRef]
44. Environment and Climate Change Canada. *Lower Mackenzie River Basin Long-Term Water Quality Monitoring Data—Canada's North*; Record ID 0177c195-13a8-4078-aa85-80b17e9e2cfe; Environment and Climate Change Canada: Gatineau, QC, Canada, 2016.
45. Sharp, H.; Rogers, Y.; Preece, J. *Interaction Design: Beyond Human-Computer Interaction*, 5th ed.; John Wiley: Indianapolis, IN, USA, 2019.
46. Gebetsroither-Geringer, E.; Stollnberger, R.; Peters-Anders, J. Interactive Spatial Web-Applications as New Means of Support for Urban Decision-Making Processes. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *4*, 59–66. [CrossRef]
47. Sun, A.Y.; Miranda, R.M.; Xu, X. Development of multi-metamodels to support surface water quality management and decision making. *Environ. Earth Sci.* **2015**, *73*, 423–434. [CrossRef]
48. Hummel, P.; Braun, M.; Augsberg, S.; Dabrock, P. Sovereignty and data sharing. *ITU J. ICT Discov.* **2018**, *25*.
49. Kukutai, T.; Taylor, J. *Indigenous Data Sovereignty: Toward an Agenda*; Anu Press: Canberra, Australia, 2016.
50. French, S.; Turoff, M. Decision Support Systems. *Commun. ACM* **2007**, *50*, 39–40. [CrossRef]
51. Geldmann, J.; Heilmann-Clausen, J.; Holm, T.E.; Levinsky, I.; Markussen, B.; Olsen, K.; Rahbek, C.; Tøttrup, A.P. What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements. *Divers. Distrib.* **2016**, *22*. [CrossRef]
52. Castillo, A.; Porro, J.; Garrido-Baserba, M.; Rosso, D.; Renzi, D.; Fatone, F.; Omez, F.V.G.; Comas, J.; Poch, M. Validation of a decision support tool for wastewater treatment selection. *J. Environ. Manag.* **2016**, *184*, 409–418. [CrossRef]
53. Harvey, F.; Chrisman, N. Boundary Objects and the Social Construction of GIS Technology. *Environ. Plan. A Econ. Space* **1998**, *30*, 1683–1694. [CrossRef]